

**Topik Mining dan Visualisasi Dinamis Dengan Metode *Latent*
Dirichlet Allocation Dalam Surat Al Baqorah**

TUGAS AKHIR

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang



RAHMAT NORIS ADITYABRILLIAN

201610370311120

Data Science

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MALANG

2020

LEMBAR PERSETUJUAN

**Topik Mining dan Visualisasi Dinamins Dengan Metode *Latent Dirichlet*
Allocation Dalam Surat Al Baqorah**

TUGAS AKHIR

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Muhammadiyah Malang**

Malang, 25 Maret 2020

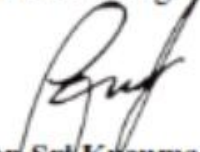
Menyetujui

Pembimbing I



Nur Hayatin, S.ST, M.Kom
NIDN.0726038402

Pembimbing II



Christian Sri Kusuma Aditya
M.Kom.
NIDN.0727029101

LEMBAR PENGESAHAN

**Topik Mining dan Visualisasi Dinamins Dengan Metode Latent
Dirichlet Allocation Dalam Surat Al Baqorah**

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

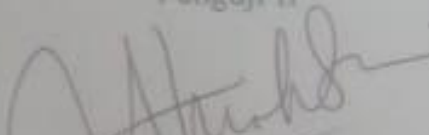
Disusun Oleh :
Rahmat Noris Adityabrillian
201610370311120

Tugas Akhir Ini Telah Diuji Dan Dinyatakan Lulus Melalui Sidang Majelis
Penguji Pada Tanggal 28 Maret 2020
Malang, 25 Maret 2020
Menyetujui

Penguji I


Denar Regata Akbi, S.Kom,
M.Kom
NIDN.0701058601

Penguji II


Zamah Sari, S.T, M.T
NIDN. 0708087701

Mengetahui,

Ketua Jurusan Teknik Informatika



Gita Indah Marthasari, ST., M.Kom.
NIP.108.0611.0442

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Rahmat Noris
Adityabrilliant
Nim : 201610370311120
Fakultas / Jurusan : Teknik / Teknik
Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Topik Mining dan Visualisasi Dinamins Dengan Metode Latent Dirichlet Allocation Dalam Surat Al Baqorah”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Malang, 3 Maret 2020

Yang Membuat Pernyataan



Rahmat Noris Adityabrilliant

Mengetahui,

Dosen Pembimbing I

Nur Havatin, S.ST, M.Kom
NIDN.0726038402

Mengetahui,

Dosen Pembimbing II

Christian Sri Kusuma Aditya,
M.Kom,
NIDN.0727029101

ABSTRAK

Pada akhir zaman ini para non-muslim yang berada di eropa sedang berbondong-bondong untuk masuk islam sehingga banyaknya muallaf yang popularitasnya terus meningkat. Hal ini menyebabkan meningkatnya pembelajaran al-quran terutama pada surat al-baqarah karena banyaknya hukum dan teknis dalam ibadah, masalahnya pembelajaran al-quran sangat susah dan terbatas akibatnya para muallaf di eropa susah untuk memahami isi dalam surat al-baqarah. Untuk menyelesaikan masalah ini dibutuhkan sebuah sistem untuk memberikan makna-makna atau topik secara ringkas dalam surat al-baqarah. Dengan dibuatnya sistem seperti ini dapat memudahkan muallaf untuk belajar atau memahami isi surat al-baqarah. Maka dari masalah ini penelitian ini dilakukan untuk mengekstraksikan topik dalam surat al-baqarah dengan menggunakan model LDA. Untuk meningkatkan kualitas topik dilakukan hyperparameter optimization dengan jumlah topik 32 dan nilai alpha 90 yang meningkatkan nilai coherence sebesar 86%. Hasil topik dari LDA akan di uji kevalidasiannya dengan pakar ahli tafsir. Dengan memberikan kuisioner dimana isi kuisioner tersebut mengambil 20% dari data training, kemudian hasil kuisioner tersebut dimasukkan kedalam *confussion matrix*. Didapatkan hasil dari *confussion matrix* nilai *recall* sebesar *confussion matrix*. Didapatkan hasil dari *confussion matrix*. Dari *confussion matrix* juga kita membandingkan jumlah topik 14 dan 32 karena kerapatan irisannya. Terbukti jumlah topik.32 lebih baik karena nilai recall sebesar 88% dibandingkan jumlah topik 14 yang 56% walaupun bubble graphnya sedikit yang menumpuk.

Kata Kunci : *topik modeling, latent dirichlect allocation, coherence, dynamic visualization, python.*

ABSTRACT

At the end of this era non-Muslims residing in Europe were flocking to convert to Islam so that the number of converts whose popularity continued to increase. This causes increased learning of the Koran, especially in Surat al-Baqarah because of the many laws and technicalities in worship, the problem is learning the Koran is very difficult and limited as a result, converts in Europe find it difficult to understand the contents in Surah Al-Baqarah. To solve this problem a system is needed to provide meanings or topics in a concise manner in Surah Al-Baqarah. With the creation of a system like this can make it easier for converts to learn or understand the contents of Surat al-Baqarah. So from this problem this study was conducted to extract the topik in Surat al-Baqarah using the LDA model. To improve the quality of the topik, hyperparameter optimization was done with the number of topik 32 and alpha value 90 which increased the coherence value by 86%. the results of the topik from the LDA will be tested for validation by expert interpreters. By giving a questionnaire where the contents of the questionnaire took 20% of the training data, then the results of the questionnaire were entered into the confusion matrix. Obtained results from the confusion matrix recall confusion matrix. From the confusion matrix we also compare the number of topics 14 and 32 because of the density of the slices. It is proven that the number of topics.32 is better because the recall value is 88% compared to the number of topics 14 which is 56% even though the bubble graph is a little piling up.

Keywords : *topik modeling, latent dirichlet allocation, coherence, dynamic visualization, python.*

LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karuni-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Banyak pihak yang terlibat dalam proses penulisan Tugas Akhir ini, maka dari itu pada kesempatan ini penulis mengucapkan terima kasih kepada seluruh pihak yang terlibat, khususnya untuk:

1. Bapak Dr. Fauzan, M.Pd., selaku Rektor Universitas Muhammadiyah Malang.
2. Bapak Dr. Ahmad Mubin, M.T., selaku Dekan Fakultas Teknik Universitas Muhammadiyah Malang.
3. Ibu Hj. Gita Indah Marthasari, ST., M.Kom., selaku ketua jurusan program studi teknik informatika Universitas Muhammadiyah Malang.
4. Ibu Nur Hayatin, S. ST., selaku dosen pembimbing I yang telah meluangkan waktu untuk memberikan arahan kepada penulis.
5. Bapak Christian Sri Kusuma Aditya, M.Kom. selaku dosen pembimbing II yang telah membimbing selama proses pengerjaan tugas akhir.
6. Bapak Christian Sri Kusuma Aditya, M.Kom, Bapak Didih Rizki Chandranegara, M.Kom dan Ibu Vinna Rahmayanti SN, S.Si, M.Si selaku dosen penguji tugas akhir yang telah memberikan saran dan kritik untuk penulis.
7. Seluruh Dosen beserta Staf Program Studi Teknik Informatika Universitas Muhammadiyah Malang yang telah memberikan ilmu dan wawasan terhadap penulis selama melaksanakan studi.
8. Seluruh keluarga penulis, khususnya kedua Orang Tua penulis, Bapak Risman Haji Guntoro dan Ibu Novi Orbandini yang telah memberikan do'a, ridho, kasih sayang dan support penuh terhadap penulis dalam menyelesaikan studi.
9. Sahabat-sahabat saya, Sirril, Abdul, Adam, Lidya, Azis, Virman, Dayat, Riki, dan Faiz. Terimakasih atas kepercayaan, waktu, dan dukungannya yang telah diberikan kepada penulis.
10. Semua anggota kelas IT-C angkatan 2016, terimakasih sudah menjadi rekan satu kelas selama kurang lebih 4 tahun

KATA PENGANTAR



Dengan memanjatkan puji syukur kehadiran Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul

“TOPIK MINING DAN VISUALISASI DINAMIS DENGAN METODE LATENT DIRICHLET ALLOCATION DALAM SURAT AL BAQORAH”

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi penentuan topik dalam surat al-baqarah, ekstraksi topik dengan LDA, optimasi coherence score. Tugas akhir ini merupakan salah satu syarat studi yang harus ditempuh oleh seluruh mahasiswa Universitas Muhammadiyah Malang, guna menyelesaikan akhir studi pada jenjang program Strata I.

Penulis menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu dan pengetahuan.

Malang, 29 Desember 2019

Penulis

DAFTAR ISI

HALAMAN JUDUL	
LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
ABSTRAK	iv
ABSTRAC	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1. 1. Latar Belakang	1
1. 2. Rumusan Masalah	2
1. 3. Tujuan Penelitian	3
1. 4. Batasan Masalah	3
BAB II LANDASAN TEORI	4
2.1. Penelitian Terkait	4
2.2. Surat Al-Baqarah	5
2.3. Machine Learning	5
2.4. Topik Modelling	6
2.5. <i>Preprocessing</i>	7
2.6. Ekstraksi Topik <i>Latent Dirichlet Allocation (LDA)</i>	8
2.7. Penentuan Jumlah Topik (Parameter Setting)	10
2.7.1. C UCI Measure	10
2.7.2. C NPMI Measure	10
2.7.3. CV Measure	11
2.8. Visualisasi Topik Model LDA	11
2.9. Pengujian	11
2.9.1. Hyperparameter Optimization	11
2.9.2. Confussion Matrix	12
BAB III METODE PENELITIAN	14
3.1. Analisa Masalah	14

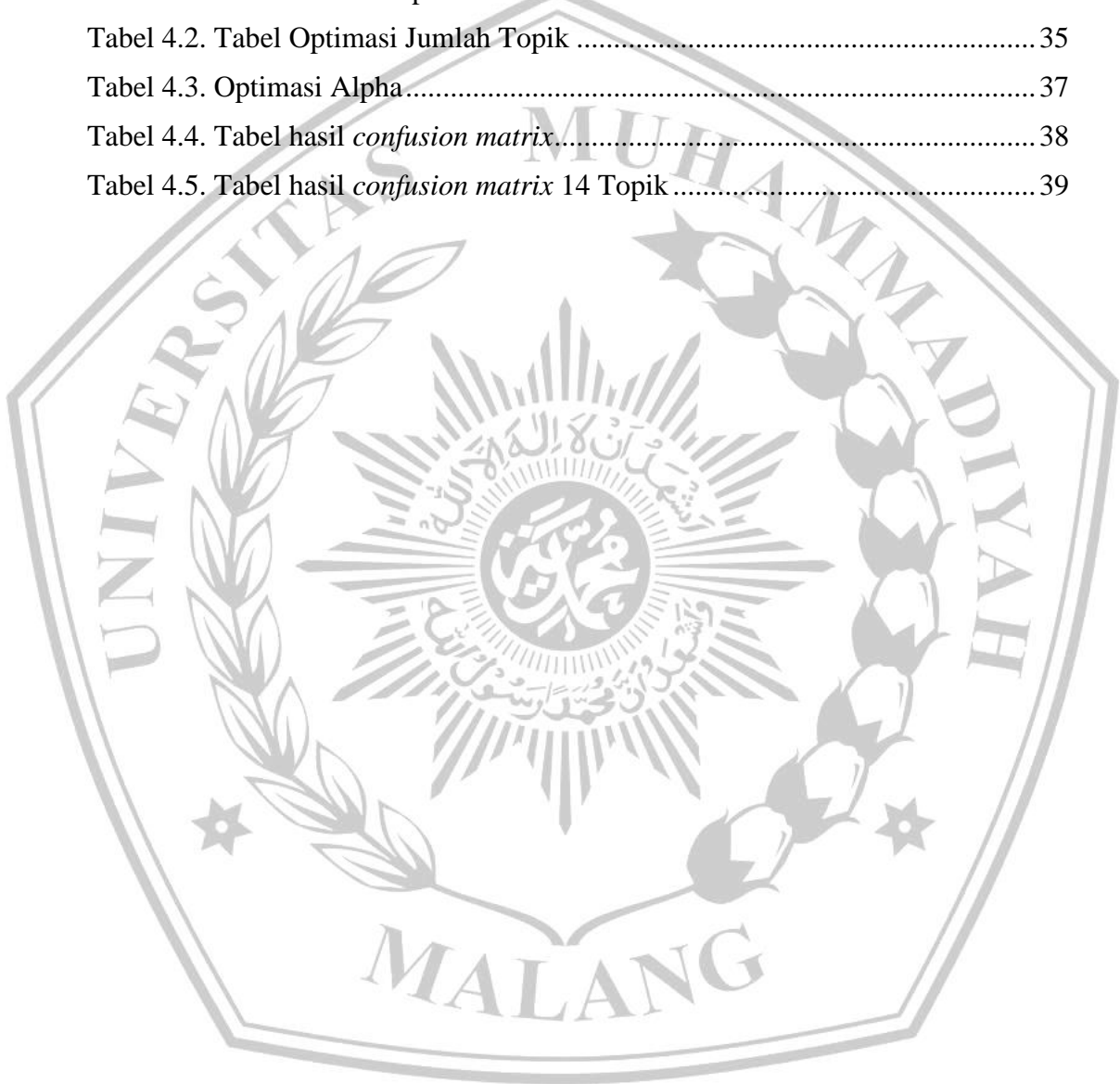
3.2.	Analisa dan Pengumpulan Data.....	14
3.3.	Analisis Perancangan Sistem.....	15
3.3.1	Tahap Preprocessing.....	16
3.3.2	Ekstraksi Topik.....	17
3.3.2.1	Latent Dirichlet Allocation	17
3.3.3	Penentuan Jumlah Topik (Parameter Setting).....	18
3.3.3.1	Coherence Score.....	19
3.3.4	Visualisasi Topik.....	21
3.4.	Pengujian Dan Analisa Hasil Program	21
3.4.1	Mallet: Machine Learning for Language Toolkit	22
3.4.2	Alpha Prior Optimization (α)	22
BAB IV	IMPLEMENTASI DAN PENGUJIAN	23
4.1.	Spesifikasi Kebutuhan <i>Software</i> dan <i>Hardware</i>	23
4.2.	Implementasi Sistem	23
4.2.1.	Implementasi Preprosesing Data.....	23
4.2.2.	Impelementasi LDA.....	25
4.2.3.	Perhitungan Coherence Score	27
4.2.4.	Visualisasi Topik.....	28
4.3.	Pengujian Sistem	34
4.3.1.	Skenario pengujian I	35
4.3.2.	Skenario pengujian II	36
4.3.3.	Skenario pengujian III.....	37
4.3.4.	Skenario pengujian IV.....	38
BAB V	PENUTUP	41
5.1.	Kesimpulan.....	41
5.2.	Saran.....	41
DAFTAR PUSTAKA.....		42
LAMPIRAN		44

DAFTAR GAMBAR

Gambar 2.1. LDA Model Diagram.....	8
Gambar 3.1. Contoh dataset instance 1 – 15.....	15
Gambar 3.2. Flowchart identifikasi dan visualisasi topik.....	15
Gambar 3.3. Tahap preprocessing	16
Gambar 3.4. Coherence score pipeline	19
Gambar 3.5. Hyperparameter Optimization Proses	21
Gambar 4.1. <i>Source code</i> untuk case folding, remove punctuation, tokenizing....	24
Gambar 4.2. <i>Source code</i> untuk <i>stopword removal</i>	24
Gambar 4.3. <i>Source code</i> untuk <i>Lemmatization</i>	24
Gambar 4.4. <i>Source code</i> untuk Dictionary dan Corpus	25
Gambar 4.5. <i>Source code</i> LDA	25
Gambar 4.6. <i>Source code</i> Ekstraksi topik LDA model gensim.....	26
Gambar 4.7. <i>Source Code</i> menghitung coherence Score	27
Gambar 4.8. <i>Source code</i> LDA Mallet.....	27
Gambar 4.9. Visualisasi Bubble Graph	30
Gambar 4.10. Contoh Memilih Topik 11	31
Gambar 4.11. Kosakata Dalam Topik 11	32
Gambar 4.12. Kontribusi Kosakata “world”.....	33
Gambar 4.13. <i>Source code</i> kontribusi tertinggi topik di setiap dokumen.....	33
Gambar 4.14. Tabel Distribusi Probabilitas Topik dalam Dokumen	33
Gambar 4.15. <i>Graph</i> jumlah topik optimal.....	35
Gambar 4.16. <i>Output Graph</i> jumlah topik optimal	35
Gambar 4.17. Visualisasi Bubble Graph 14 Topik.....	35

DAFTAR TABEL

Tabel 2.1. Notasi LDA model diagram	8
Tabel 2.2. <i>Confussion Matrix</i>	12
Tabel 3.1. Proses text preprocessing	17
Tabel 3.2. Contoh coherence pada data	20
Tabel 4.1. Contoh Hasil Topik-Kosakata	28
Tabel 4.2. Tabel Optimasi Jumlah Topik	35
Tabel 4.3. Optimasi Alpha	37
Tabel 4.4. Tabel hasil <i>confusion matrix</i>	38
Tabel 4.5. Tabel hasil <i>confusion matrix</i> 14 Topik	39



DAFTAR PUSTAKA

- [1] M. A. Siddiqui, "Discovering the Thematic Structure of the Quran using Probabilistic Topic Model," no. May 2015, 2013, doi: 10.1109/NOORIC.2013.55.
- [2] H. Kettani, "Muslim Population in Europe : 1950 – 2020," vol. 1, no. 2, pp. 154–164, 2020.
- [3] M. Aderi, C. Noh, and A. H. Tamuri, "The Study of Quranic Teaching and Learning : United Kingdom Experience," vol. 5, no. 16, pp. 313–317, 2014, doi: 10.5901/mjss.2014.v5n16p313.
- [4] N. Ali and M. Ahmed, "Expansion Strategy in the Translation of Iltifāt in Sūrat Al-Baqarah," no. 5, pp. 54–65, 2016.
- [5] A. I. Technology *et al.*, "Knowledge exploration: Selected works on Quran ontology development," pp. 8–11, 2017, doi: 10.1007/s10579-012-9186-z.
- [6] D. B. Nugraheni, M. A. Bijaksana, and E. Darmawiyanto, "Analisis Dan Implementasi Pencarian Kata Berbasis Konkordansi Dan N-Gram Pada Terjemahan Al-Quran Berbahasa Indonesia Analysis and Implementation Concordance Search and N-Gram for Words in Al-Quran English Translation," vol. 4, no. 3, pp. 4713–4718, 2017.
- [7] M. I. Rahman, N. A. Samsudin, A. Mustapha, and A. Abdullahi, "Comparative analysis for topic classification in Juz Al-Baqarah," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 406–411, 2018, doi: 10.11591/ijeecs.v12.i1.pp406-411.
- [8] A. Ta and S. Sudin, "Exploring Meaningful Concepts of Al-Baqarah Chapter Using Text Mining Approach," no. July, pp. 25–27, 2018.
- [9] D. Rolliawati, I. S. Rozas, and M. Ratodi, "Text Mining Approach for Topic Modeling of Corpus Al Qur ' an in Indonesian Translation."
- [10] M. H. Panju, "Statistical Extraction and Visualization of Topics in the Qur ' an Corpus by," 2014.
- [11] M. Alhawarat, "Extracting Topics from the Holy Quran Using Generative Models," no. December 2015, 2016, doi: 10.14569/IJACSA.2015.061238.
- [12] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring Topic Coherence over many models and many topics," no. July, pp. 952–

961, 2012.

- [13] M. Röder and A. Hinneburg, “Exploring the Space of Topic Coherence Measures.”
- [14] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, “Automatic Evaluation of Topic Coherence,” no. June, pp. 100–108, 2010.
- [15] N. Aletras and R. Court, “Evaluating Topic Coherence Using Distributional Semantics,” no. 2009, 2011.
- [16] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” vol. 3, pp. 993–1022, 2003.
- [17] A. El Mouatasim and O. Jaouad, “Topic Classification of Arabic Text in Quran by using Matlab,” vol. 53, no. November 2018, pp. 215–224, 2019, doi: 10.1007/978-3-030-12048-1.
- [18] T. Xie, P. Qin, and L. Zhu, “Study on the Topic Mining and Dynamic Visualization in View of LDA Model,” *Mod. Appl. Sci.*, vol. 13, no. 1, p. 204, 2018, doi: 10.5539/mas.v13n1p204.
- [19] H. M. Wallach, D. Mimno, and A. McCallum, “Rethinking LDA: Why priors matter,” *Adv. Neural Inf. Process. Syst. 22 - Proc. 2009 Conf.*, pp. 1973–1981, 2009.
- [20] I. Science, “University of Arkansas at Little Rock Mallet vs GenSim: Topic modeling for 20 news groups report .,” 2016.
- [21] J. Boyd-Graber, Y. Hu, and D. Mimno, *Applications of topic models*, vol. 11, no. 2–3. 2017.
- [22] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, and M. Welling, “Fast collapsed gibbs sampling for latent dirichlet allocation,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 569–577, 2008, doi: 10.1145/1401890.1401960.



UNIVERSITAS MUHAMMADIYAH MALANG
FAKULTAS TEKNIK
PROGRAM STUDI TEKNIK INFORMATIKA

Jl. Raya Tlogomas 246 Malang 65144 Telp. 0341 - 464318 Ext. 247, Fax. 0341 - 460782

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

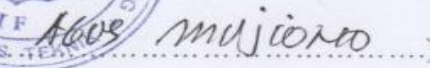
Nama Mahasiswa : *Rahmat Floris Aditya Brillian*
 NIM : *201610370311120*
 Judul TA : *Topic mining dan visualisasi dinamis dengan metode Latent Dirichlet Allocation dalam surat Al-Baqorah.*

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	7 %
2.	Bab 2 – Daftar Pustaka	25 %	23 %
3.	Bab 3 – Analisis dan Perancangan	25 %	24 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	14 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	4 %
6.	Makalah Tugas Akhir	20%	19 %

Mengetahui,

Dosen Pembimbing/Tim Cek Plagiasi



 (A. M. Mujiono)

*) Hasil cek plagiarisme bisa diisikkan oleh salah satu pembimbing